A Work Project, presented as part of the requirements for the Award of a Master Degree in Economics from the NOVA – School of Business and Economics.

Education and Wage Inequality in Portugal

Henrique Robles Machado Duarte Neves,

Student Number 29313

A Project carried out on the Master in Economics Program, under the supervision of:

Professor Pedro Portugal

04-01-2021

Education and Wage Inequality in Portugal

Abstract

The main goal of this Master's thesis is to identify the impact of education on changes in the Portuguese wage distribution using quantile regression methods. I explore the counterfactual decomposition method proposed by Machado and Mata (2005) to decompose changes in the wage distribution from 1994 to 2018 into the factors contributing to those changes: composition effects (characteristics of the Portuguese working force) and structural effects (returns to those characteristics). The results show that education was a key factor contributing to higher wages but also greater wage inequality.

Keywords: Wage inequality, Returns to education, Quantile regressions, Composition effects, Structural effects, Counterfactual analysis

---

Firstly, I would like to thank my thesis supervisor, Professor Pedro Portugal, for always being there to help me and guide me for the last few months. I am very grateful for all the empirical knowledge and ideas shared by the Professor, they were crucial for the pursuance of this thesis. It was really a honour to work and learn with you. I would also like to thank Instituto Nacional de Estatística (INE) for giving me access to the data set used in this work (*Quadros de Pessoal*) and to Professor Pedro Raposo, for kindly helping me with the Stata code used in this work. I am also thankful to all the people I encountered at Nova: the professors for providing an education of excellence; and all the amazing friends I made in the last five years, it has been an incredible journey. Finally, a special thanks to my family for all the unconditional support, patience, and trust, you are all a huge inspiration to me and certainly represent a huge part of this work.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

This paper studies the impact of education on wage inequality in Portugal, from 1994 to 2018. Using the methodology proposed by Machado and Mata (2005), I decompose the changes that occurred in the Portuguese wage distribution in this period into two factors: changes in the productive characteristics of workers and changes in the return to labour market skills. The results show that education contributed unequivocally to higher wages and to greater wage inequality, despite the observed decline in inequality measures during the period.

This finding goes in hand with the previously addressed inequality increasing effect of education reported by an extensive list of studies employing quantile regression methods. Despite the fact that educational policies are widely seen as tools to mitigate income inequalities, the evidence obtained in this work suggests that the benefits of education are not equally shared by workers at different positions in the wage distribution: education has a greater impact on the wages of workers at the top of the wage distribution than it has on workers at the bottom.

The paper is structured as follows: Section 2 reviews the literature around decomposition methods and wage inequality. Section 3 presents the methodology used in this paper and Section 4 characterizes the data set used in this work. Section 5 describes the Portuguese wage distribution, separating the role of the characteristics of the working force and the Portuguese wage structure. Section 6 presents the results from the Machado and Mata (2005) decomposition method and Section 7 discusses the implications of the reported results and concludes.

2 Literature Review

The main ingredient of this thesis is the methodology proposed by Machado and Mata (2005). This decomposition method resembles the spirit of the conventional Oaxaca-Blinder method that decomposes changes in mean wages between two groups into a composition effect (changes in worker's characteristics) and a wage structure effect (changes in regression coefficients). However, it has the additional feature of extending the Oaxaca-Blinder decomposition effects on mean wages to the entire wage distribution.

The decomposition of differences in wage distributions of two groups depends on the construction of a counterfactual wage distribution, obtained by manipulating wage-setting functions through counterfactual distributions of observed characteristics of the working force. To achieve this, Machado and Mata rely on quantile regression methods proposed by Koenker and Basset (1978). The authors estimate conditional wage distributions by quantile regressions and use resampling procedures to estimate marginal densities of wages consistent with these conditional wage distributions and with hypothetical distributions of worker's attributes. For example, one could estimate the marginal density of wages in 2018 as if all covariates (or just one) were distributed as in 1994. Comparing these counterfactuals with the actual marginal distribution of wages in 2018 would give us the effect of changes in characteristics of workers on changes in the distribution of wages (composition effect or individual effect of each variable on changes in the distribution of wages). By applying this methodology to Portugal, using data between 1986 and 1995, the authors found that the increase in educational levels during the period was at the root of the observed increase in wage inequality.

This decomposition method has already been used in the Portuguese labour market literature. Cardoso, Guimarães, Portugal and Raposo (2016) use the Machado and Mata procedure to identify the sources of gender differences in the distribution of wages in Portugal for the years 1991 and 2013 and to compare how the sources of variation have changed between the two years. The authors find a decrease in the magnitude of the gender gap from 1991 to 2013 and a higher role of structural effects in explaining the gender gap in both years. Portugal, Raposo and Reis (2018) also use this method to disentangle the structural and compositional changes in the distribution of wages in Portugal from 1988 to 2013 and to isolate the effects of changes in educational levels for workers and observed changes in the returns to labour market experience, finding that education was the most decisive factor in shaping the Portuguese wage distribution.

Many other methodological papers about decomposition methods in economics were written since the emergence of the seminal papers of Oaxaca (1973) and Blinder (1973), that compose the Oaxaca-Blinder decomposition method. A common goal of these recent methods is expanding Oaxaca-Blinder decomposition to other distributional statistics beyond the mean to understand the underlying factors behind inequality growth since the late 1970s. Fortin, Lemieux and Firpo (2011) provided an extensive overview of the literature around many of these decomposition methods, from which I highlight Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996) and as already referred Machado and Mata (2005) methodologies.

Juhn, Murphy, and Pierce (1993) proposed a residual imputation procedure that tried to explain the rising wage inequality in the US during the 1970s and 1980s. The authors developed a framework using OLS wage regressions that decomposed changes in inequality into three sources: composition effects, structural effects and changes in the distribution of residuals. By simulating counterfactual distributions, this method is able to identify the contribution of changing "quantities", the marginal contribution of changing "prices" and the marginal contribution of changing residuals to changes in inequality (where the residuals represent unobserved ability and its respective market price). The author's main conclusion was that the trend towards greater wage inequality beginning in the 1970s was attributable primarily to increases in the price of both unobserved and observed dimensions of skill (especially unobserved). They attributed this to a pervasive demand shift toward the most skilled, which was later interpreted as being driven by Skill-Bias Technological Change.

DiNardo, Fortin and Lemieux (1996) proposed an alternative explanation focused on the role of wage-setting institutions in explaining growth in inequality in the US between 1979 and 1988, by accounting for the effects of two possible institutional factors: de-unionisation of workers and fall in real minimum wages. The authors proposed a semi-parametric procedure to estimate and analyse counterfactual densities of wages, by applying kernel density estimates to appropriately weighted samples using a reweighting factor. The main idea consists in estimating simple counterfactual densities, such as the hypothetical density of wages in 1989 if worker attributes (for example their union status) had remained at their 1979 levels. This is done by using a reweighting factor that replaces the marginal density of a certain covariate in 1989 with the marginal density of the same covariate in 1979 (Lemieux (2002) extended this methodology to account for the role of prices, previously referred as "unexplained changes"). The main conclusion of the authors was that labour market institutions (especially the role of unions and the minimum wage) were as important as supply and demand factors to explain changes in the US distribution of wages.

The three methodologies described above (Juhn, Murphy and Pierce (1993); DiNardo, Fortin and Lemieux (1996); Machado and Mata (2005)) are considered to be among the most widely used decomposition methods in the labour economics literature and set the theme for the emergence of many influential ideas about wage inequality, some of which I will describe.

Firstly, Lemieux (2005) (thereafter Lemieux (2006b)) documented the marked growth in residual wage inequality in the US between 1973 and 2003. This development had already been documented by Juhn, Murphy and Pierce (1993), that considered that this steep growth accounted for much of the growth in overall wage inequality. The main difference between the two papers was the source attributed to the growth in residual inequality. Contrary to Juhn, Murphy and Pierce, Lemieux did not focus solely on the skill-premium hypothesis. The author also considered the hypothesis that the dispersion in unobserved skills might had been increasing due to composition effects. To Lemieux, growth in residual wage inequality was essentially originated by the secular increase in experience and education, two factors associated with higher within-group wage dispersion, as previously explored in the literature: 1) More experienced workers have more dispersed distributions of skills due to differential investments in on-the-job training, as proposed by Mincer (1974); 2) residual variance of wages increases with schooling levels as proposed by Mincer (1997). Hence, if the share of groups in the economy having more dispersed wages increases (via composition effects) residual inequality might be increasing.

This idea is also explored in Machado and Mata (2005) to explain the inequality increasing effect of education. In an economy composed of two types of workers- low-and high-skilled (measured by years of schooling)- an increase in the proportion of high-skilled workers is conventionally perceived to reduce wage inequality via two mechanisms: a price effect (relative wage for skilled workers decreases) and a composition effect (proportion of skilled workers earning higher wages increases). The problem with this reasoning is that it neglects wage variation within groups. An increase in the proportion of skilled workers in the economy could be inequality increasing if there is more wage dispersion in this group. The extent of within-group inequality can be measured using quantile regressions, by comparing regression coefficients for different quantiles. As the authors point out, for Portugal, between 1986 and 1995, the effect of education was more decisive at the top quantiles of the wage distribution than at the bottom, meaning that education contributed to more wage dispersion, and hence that better-educated workers exhibited more dispersed wage distributions than lesser-educated workers. Additionally, the returns to education increased at the top of the wage distribution and remained constant at the lowest quantiles. Campos and Reis (2017) also document an increase in the return to education along the Portuguese wage distribution in 2013. By comparing the returns to education at different quantiles (90/10, 90/50, 50/10) between 1986 and 2013, the authors provide different dispersion measures that exhibit increasing inequality in the returns to education.

The second idea refers to Autor, Katz, and Kearney (2005). The authors used quantile regressions to explain the existing divergence in inequality trends in the US since 1987: steady rise in upper-tail inequality against flat or declining lower-tail inequality. By using an extension of Machado and Mata (2005) quantile approach that nests both Juhn, Murphy and Pierce (1993) and DiNardo, Fortin and Lemieux (1996) methods, the authors concluded that the mechanical effects of shifts in labour force composition (proposed by Lemieux (2005) did not explain the asymmetric growth of wage inequality. As for the sources of the divergence in upper-and lower-tail inequality, Goos and Manning (2007) would later propose an eloquent explanation for the UK wage distribution that was afterwards applied to the US by Autor, Katz, and Kearney (2008): the impact of computerization on changing demand for job tasks. The main idea is that computerization led to structural shifts in the demand for job tasks compatible with the "polarization of work"- sharp rise in demand for "abstract tasks" (that require higher educational attainments); reducing demand for "routine tasks" performed by plenty middle educated workers, with little impact on "manual tasks" used by many low-skilled workers in service jobs.

3 Empirical Methodology

3.1 Quantile Regression Methods (QR)

The first step to estimate quantile regression methods is the conditional quantile function (CQF). The CQF for a random variable Y continuously distributed, given a set of covariates x is represented by:

$$Q_\theta(y_i|x_i) = F_Y^{-1}(\theta|x_i) \tag{1}$$

where $F_Y(y|x)$ is the conditional distribution function of Y given x, and θ represents the quantile of the conditional distribution. The dependent variable, y_i refers to the logarithm of hourly real wages (or w_i) and x_i corresponds to the set of worker and firm characteristics: gender, age, age squared, tenure, tenure squared and education (years of schooling) for workers and firm size and industry for the firms. Hence, the conditional quantile model is represented by:

$$Q_\theta(w|x) = x^{'}\beta(\theta) \tag{2}$$

where $\beta(\theta)$ is a vector of coefficients. Following Koenker and Basset (1978) we can estimate $\beta(\theta)$ by minimizing the following expression in β:

$$\begin{aligned} \beta(\theta)^{'} &= argmin \sum_{i=1}^{N} \rho_\theta(w_i - x_i\beta) \\ &= \sum_{(i:w_i \geq x_i\beta_\theta)}^{N} \theta|w_i - x_i\beta_\theta| + \sum_{(i:w_i < x_i\beta_\theta)}^{N} (1-\theta)|w_i - x_i\beta_\theta| \end{aligned} \tag{3}$$

where $\rho_\theta(u)$ (also called check function) is an asymmetric loss function.

Quantile regression methods allow some heterogeneity in the coefficients, by taking into account the position of an individual along the wage distribution and running a separate regression for each value of θ. Hence, $\beta(\theta)$ will correspond to the return of the different worker and firm attributes on the θth quantile of the wage distribution, conditioning on a set of covariates.

The main motivation to apply quantile regressions is that they provide information about the whole conditional wage distribution. This is especially relevant when covariates affect other moments of the distribution of wages apart from the mean, for example in the form of heteroskedasticity. Additionally, by comparing coefficients at different points of the distribution (for example upper and lower deciles) we can obtain measures of inequality, and understand which covariates contributed to more/less inequality. If coefficients are not constant across quantiles, there may be signs of within-group inequality: workers with the same characteristics but located at different points of the wage distribution having different rates of return.

3.2 Machado and Mata (2005) Decomposition Method

The conditional quantile function $Q_\theta(w|x)$ fully characterizes the conditional distribution of wages for $\theta \in (0,1)$. Machado and Mata propose a resampling procedure that simulates a random sample from the conditional distribution of wages that is consistent with the QR model.

The main idea of this procedure rests in the probability integral transformation theorem: By randomly drawing $\theta_1, \theta_2, ..., \theta_m$ from a uniform distribution (0,1), and for each θ_i estimating the model in (2) and obtaining m vectors $\hat{\beta}(\theta_i)$, we can ensure that for a given value of covariates x_i, the conditional quantiles function of wages evaluated at x_i , $\{x_i' \hat{\beta}(\theta_i)\}_{i=1}^m$ will be a random sample from the estimated conditional distribution function of wages given x_i. Hence, this simulation technique provides a full characterization of the conditional distribution of wages.

To obtain a marginal density of wages, the authors propose integrating out the conditioning covariates. This integration is performed with respect to different joint distributions of covariates, $g(x)$ at time t ($t = 0$ refers to 1994 and $t = 1$ to 2018). The procedure is summarized as follows:

For the data set at time t :

1. Generate $\theta_i, i = 1, 2; ..., m$ from a U [0,1] and estimate $\hat{\beta}_t(\theta_i)$ for each t and i;

2. Generate a random sample of size m with replacement from the rows of $g(x)$ denoted by $\{x_i^*\}, i = 1, .., m$;

3. Obtain $w_i^*(t) = \{x_i^*(t)' \hat{\beta}_t(\theta_i)\}_{i=1}^m = g(x_i^{*'} \hat{\beta}_t(\theta_i))$.

This constitutes a random sample of size m from the marginal distribution of wages.

Alternatively, instead of drawing random values for the quantiles from a uniform distribution, one could estimate 99 quantile regressions, one from each percentile from 1 to 99. In this case, instead of drawing the rows of $g(x)$ from replacement, we could directly apply a random sample from the distribution of covariates $g(x)$ to the transposed vector of 99 coefficients $\beta_t(\theta_i)$. In other words, by attributing a random sample from $g(x)$ to each percentile we would obtain a draw from the estimated conditional distribution of wages.

In this setting, $g(x)$ may represent the actual distribution of covariates, but also any distribution of interest. For example, when $g(x)$ represents the distribution of covariates in 1994, by applying it to the coefficients matrix $\beta_0(\theta_i)$ we obtain a sample from the conditional distribution of wages in 1994 : $f^*(w(0))$. If, on the other hand, we wanted to construct a counterfactual distribution, we could apply the distribution of covariates in 1994 ($g(x(0))$) to the coefficients matrix in 2018 ($\beta_1(\theta_i)$). This would give us the counterfactual wage distribution in 2018 if covariates were distributed as in 1994 ($f^*(w(1); g(x(0)))$). By comparing:

- $f^*(w(1))$ with $f^*(w(1); g(x(0)))$ we obtain the composition effect: the effect of covariates on changes in the wage distribution.
- $f^*(w(1); g(x(0)))$ with $f^*(w(0))$ we obtain the structural effect: the effect of QR coefficients on changes in the wage distribution.

In this way, it is possible to decompose changes in the wage distribution into changes in covariates and coefficients. This method applies the traditional Oaxaca-Blinder decomposition to the entire wage distribution, rather than just the mean. Additionally, it allows us to construct counterfactual distributions, and isolate the impact of individual covariates and their coefficients on changes in the wage distribution.

The impact of a single covariate y is obtained by simulating the wage distribution that would occur if only one covariate was distributed as in 1994 and the remaining covariates as in 2018 ($f^*(w(1); y(0))$). This would give us the counterfactual wage distribution in 2018 if, for example, education was distributed as in 1994. By comparing it with the actual distribution of wages in 2018, we obtain the education composition effect. This implies reallocating the weights given for the different classes that compose a specific covariate. For example, for the variable gender, this would imply changing the proportion of men and women in the workforce in 2018 to the levels observed in 1994. By the same token, we could also identify the impact of changing returns to each covariate, by computing the difference between the counterfactual wage distribution in 1994 if only the coefficient under scrutiny was distributed as in 2018 ($f^*(w(0); \beta_y(1))$) with the wage distribution in 1994. Summing up:

- $f^*(w(1)) - f^*(w(1); y(0))$: individual composition effect
- $f^*(w(0); \beta_y(1)) - f^*(w(0))$: individual structural effect

4 Data description

The analysis performed in this paper is based on *Quadros de Pessoal*, an annual mandatory survey conducted by the Ministry of Labour and Social Security which contains information about all establishments that employ at least one dependent employee in Portugal. The survey covers detailed information about worker's wage components (wage, hours worked), worker's attributes (such as gender, age, tenure at the firm, schooling), and firm-level characteristics (such as industry and dimension of the firm).

My data set consists of two samples of 100,000 workers for the years 1994 and 2018 respectively, randomly extracted from *Quadros de Pessoal*. The two samples restrict to full-time wage earners aged between 18 and 65 years, with base wage higher than 80% of Portuguese minimum wage in each year (lowest admissible wage for apprentices) and with information regarding gender, age, tenure, schooling, and firm size and industry of the firm. Workers from the Agriculture and Fishery sectors were excluded from the sample.

The wages are defined as the sum of all regular payments (base wage, extra hours, and other regular payments). The hourly wages are obtained by dividing regular wages by the number of hours worked (normal and extra hours). The wages are deflated by the Portuguese CPI with the base year 1994. In *Quadros de Pessoal*, the education level of an individual is defined by a categorical variable that reports the maximum education level obtained. Additionally, a continuous variable for education was created based on *Quadros de Pessoal* nomenclature, giving information about the minimum amount of years needed to attain the highest education level reported in the data set.

5 Composition and Wage Structure

5.1 Composition of the workforce

To decompose the changes that occurred in the Portuguese wage distribution from 1994 to 2018, we should disentangle the factors that contributed to those changes: changes in the productive characteristics of the workforce (covariates); and changes in the return to these characteristics (coefficients or wage structure). During this period, as Figure 1 shows, the wage distribution shifted to the right (real rise of the average hourly wage of 25%) and became less dispersed.

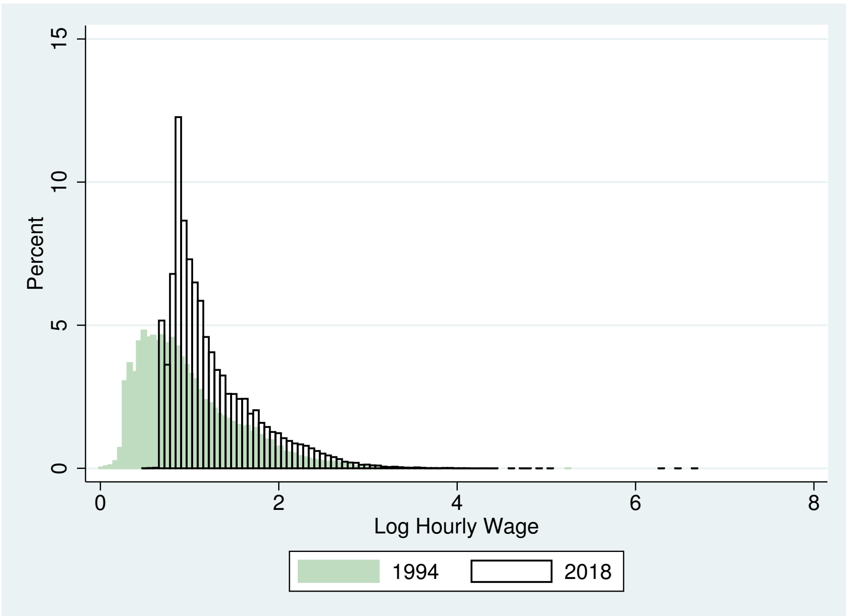


Figure 1: Log hourly wage distribution for 1994 and 2018

Table 1 reports inequality measures for 1994 and 2018. The variance and quantile ratios refer to the logarithm of real hourly wages, and the Gini Index to the real hourly wages. In line with Figure 1, all the inequality measures point o a decrease in wage inequality. Additionally, this decrease appears to be essentially concentrated in the lower half of the wage distribution: The 90/10 log wage ratio decreased by 22.4 log points, and 76% of this drop is explained by the drop in the 50/10 ratio. One possible explanation for this pattern could be the sharp increase in the minimum wages that Portugal witnessed in the period and the fact that the impact of minimum wages is concentrated at the lower tail of the wage distribution, as explored by Centeno and Novo (2009).

Inequality Measures	1994	2018
Variance	0.322	0.247
90/10	1.407	1.184
90/50	0.938	0.885
50/10	0.469	0.299
Gini	0.355	0.316

Table 1: Inequality Measures

One of the main goals of this work is to explain which factors contributed to the main changes in the wage distribution previously described. Traditional wage equations use the variables included in specification (2) as explanatory variables. For Portugal, these variables explain from 43% to 55% of the wage variation in 2018 and 1994, respectively.

Table 2 describes the main characteristics of the Portuguese workforce. From there, we can observe an increase in female labour market participation rate; an increase in the average age of workers, which is very likely related to the fact that average educational levels increased significantly, from 6.65 to 10.28 years of education; and finally a decrease in the average job tenure at the firm and the average dimension of firms.

	1994	2018
Gender (% Female)	0.39	0.46
Age	35.87	40.96
	(11.12)	(11.02)
Tenure	8.14	8.03
	(8.61)	(9.18)
Firm Size	14.65	12.43
	(115.28)	(121.36)
Education	6.65	10.28
	(3.64)	(3.36)
Educational Level (%)		
Less than 9th grade	0.67	0.21
9th grade	0.15	0.26
Secundary Education	0.13	0.31
Higher Education	0.05	0.22

Table 2: Covariates - descriptive statistics

Notes: Sample means (Standard errors in parenthesis).

Additionally, a closer examination regarding educational levels allows us to better understand the nature of composition changes in the education of workers. There is a marked increase in the proportion of workers with a high school diploma and higher education and a substantial decrease in the proportion of workers with less than the 9th grade.

As discussed in the literature review, education and age are the variables associated with higher within-group dispersion. Hence, if the workforce becomes older and more educated, these two factors could be contributing to more residual wage inequality. Following Centeno and Novo (2009), Table 6 in the Appendix reports the evolution of wage dispersion within nine narrow education and age groups. The variable age is partitioned into three classes (<36, 36-45, >45)), as is education (6 or less years of schooling, 9th grade or Secondary Education, College degree). The inequality measures in Table 6 confirm that wage dispersion increases with education and with age (for the three education-skill groups), meaning older and more educated individuals have more dispersed wage distributions. The 90/10 quantile ratio has decreased for all the nine skill groups, except for older individuals (age greater than 45) with a college degree. The lower-tail inequality decreased for all the three education groups. The upper-tail inequality only increased for the higher educated group.

5.2 Wage Structure

To characterize the wage structure, I estimate quantile regressions for the logarithm of real hourly wages using the model specification in (2). The coefficients $\beta(\theta)$ represent the return of an individual covariate at the θth quantile of the distribution of real hourly wages. Table 3 exhibits quantile regressions for $\theta = \{10, 25, 50, 75, 90\}$. There are essentially two dispersion measures we are interested in. The first is to compare, for each year, coefficients for different points of the wage distribution. The second is to compare how coefficients have changed for each quantile between the two years. Both will describe the impact of each covariate upon wage dispersion.

From Table 3, we can observe that women earn less than men at all the selected quantiles and that the gender gap increases along the wage distribution, being much more pronounced in the highest quantiles. These results imply that a regression upon the mean would not adequately capture the existing differences in the wage distributions of the two groups and that women exhibit less dispersed wage distributions than men. Women earned 22.1 log points (24.7 per cent) and 15.5 log points (16.8 per cent) less than men at the conditional median in 1994 and 2018, respectively. The gender gap became much more narrow between the two years, despite the fact this gap decreased much less at the higher quantiles.

Regarding age, a proxy for labour market experience, there are two important findings. First, the returns to age are positive and monotonically increasing with the quantiles. In other words, more experienced workers have higher wages, especially at the highest quantiles. Secondly, the returns to experience have dropped considerably between 1994 and 2018 for all quantiles. Contrarily to this pattern, the returns to tenure at the firm are more homogeneous and have increased at the upper half of the wage distribution, remaining roughly constant at the left tail of the wage distribution. The coefficients for firm size capture some of the heterogeneity in firm's pay policies: larger firms pay higher wages at all the selected quantiles. The returns are constant across the quantiles in 1994 and have dropped considerably across the entire distribution from 1994 to 2018, especially at the left tail of the wage distribution.

One of the most interesting results from Table 3 is the inequality increasing effect of education. The impact of education on wages is positive at all points of the wage distribution, but it is increasing with the quantiles. This suggests that education is more valued at higher-paid jobs. Hence, increasing the educational levels of the workforce would translate into a more dispersed distribution of wages because more educated workers exhibit more spread out wage distributions. The returns to education have decreased in all quantiles except the 90th quantile from 1994 to 2018. Nevertheless, this decrease was much more striking at the bottom of the wage distribution. This fact, associated with the monotonic increase of the impact that education has on wages along the wage distribution shows evidence that education contributed to higher wage dispersion in Portugal from 1994 to 2018.

Table 3: Quantile Regressions - 1994 and 2018

	1994					2018				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Gender (Female=1)	-0.140***	-0.177***	-0.221***	-0.261***	-0.294***	-0.064***	-0.100***	-0.155***	-0.200***	-0.254***
	(0.002)	(0.000)	(0.006)	(0.004)	(0.005)	(0.001)	(0.005)	(0.001)	(0.001)	(0.004)
Age	0.023***	0.030***	0.037***	0.046***	0.059***	0.005***	0.010***	0.021***	0.032***	0.042***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Age squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.012***	0.012***	0.011***	0.010***	0.005***	0.009***	0.012***	0.017***	0.018***	0.016***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Tenure squared	-0.000***	-0.000***	-0.000***	-0.000***	0.000	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm size (Log)	0.061***	0.065***	0.065***	0.064***	0.067***	0.021***	0.027***	0.038***	0.047***	0.049***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Education	0.039***	0.053***	0.070***	0.085***	0.095***	0.023***	0.037***	0.059***	0.081***	0.099***
(Years of Schooling)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Constant	-0.529***	-0.619***	-0.674***	-0.725***	-0.811***	0.371***	0.183***	-0.176***	-0.507***	-0.694***
	(0.021)	(0.014)	(0.007)	(0.005)	(0.033)	(0.009)	(0.002)	(0.021)	(0.007)	(0.046)

Notes: Wage Regressions-Quantile Regressions. The coefficients were obtained as of specification (2). Analytical Standard errors in parenthesis. **** p<0.01, ** p<0.05, * p<0.1

6 Changes in the distribution of wages

6.1 Aggregate Decomposition

Table 4 exhibits the results of the aggregate decomposition. Columns 1 and 2 represent the estimated conditional wage distributions in 1994 and 2018, respectively, and column 3 the differences between the two distributions. The results from our model show great resemblances to the changes in the empirical wage distribution: inequality measures decreased considerably, this decrease being more concentrated in the lower half of the wage distribution. The wage distribution has shifted to the right, but this shift was more pronounced at the lower quantiles.

Table 4: Aggregate Decomposition

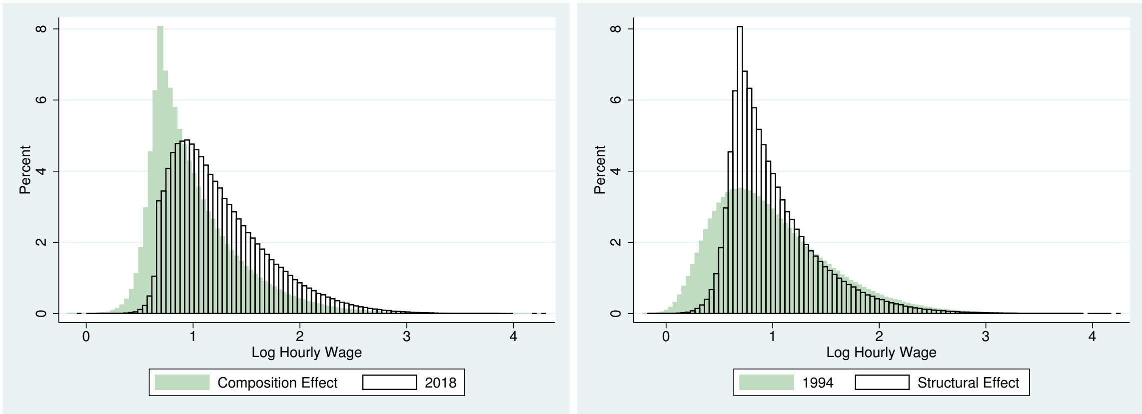
	(1) 1994	(2) 2018	(2)-(1) Change	(x1b1-x0b1) Covariates	(x0b1-x0b0) Coefficients
10th quant.	0.388***	0.762***	0.374***	0.154***	0.220***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
20th quant.	0.530***	0.860***	0.329***	0.179***	0.151***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
30th quant.	0.651***	0.950***	0.299***	0.212***	0.087***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
40th quant.	0.769***	1.043***	0.274***	0.236***	0.038***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
50th quant.	0.892***	1.145***	0.253***	0.257***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
60th quant.	1.028***	1.262***	0.234***	0.276***	-0.042***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
70th quant.	1.190***	1.405***	0.215***	0.294***	-0.079***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
80th quant.	1.400***	1.592***	0.191***	0.306***	-0.114***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90th quant.	1.715***	1.876***	0.161***	0.301***	-0.140***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
90/10	1.327***	1.115***	-0.213***	0.148***	-0.360***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
90/50	0.824***	0.731***	-0.092***	0.044***	-0.137***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
50/10	0.503***	0.383***	-0.120***	0.103***	-0.224***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: The estimated wage distributions were obtained by applying the QR coefficients to different distributions of covariates. A random sample of 10,000 individuals was attributed to each percentile from 1 to 99. The standard errors (in parenthesis) are obtained from 100 bootstrap replications. **** p<0.01, ** p<0.05, * p<0.1

Columns 4 and 5 decompose the changes reported in column 3 into a composition and a structural effect. The first one is obtained by comparing the wage distribution in 2018 (x1b1) with the counterfactual distribution in 2018 if all covariates were distributed as in 1994 (x0b1), hence isolating the impact of covariates (x's) on changes in log hourly wages. Composition effects contributed to the observed shift to the right of the whole conditional wage distribution, with the magnitude of this shift increasing monotonically with the quantiles. As a result, the observed changes in the composition of the workforce would have increased the 90/10 ratio by 14.8 log points, holding coefficients constant.

The structural effect, shown in column 5, is obtained by comparing the counterfactual wage distribution in 2018 if covariates were distributed as in 1994 (x0b1) with the wage distribution in 1994 (x0b0), hence isolating the impact of coefficients (β's) on changes in log hourly wages. The results suggest that: 1) structural effects contributed to the observed shift of the wage distribution to the right only up to the 40th quantile, reinforcing the composition effect. From the 40th quantile onwards, coefficients had a negative impact on the wage distribution, since the returns to some attributes have decreased between 1994 and 2018; 2) Contrarily to composition effects, structural effects contributed to a less unequal wage distribution: holding the composition of the workforce constant, changes in coefficients would have decreased the 90/10 ratio by 36 log points. This decrease outweighed the opposite increase due to composition effects, leading to a decline in estimated inequality measures from 1994 to 2018.

Figure 2 shows the estimated wage distributions based on 99 quantile regressions. On the left side, two distributions are presented: the wage distribution in 2018, and the counterfactual distribution in 2018 if all covariates were distributed as in 1994. As shown in Table 4, composition effects contributed to a clear shift to the right of the wage distribution. The right side, in turn, presents the wage distribution in 1994 and the counterfactual distribution in 2018 if all covariates were distributed as in 1994. As we can see, there were evident improvements of the wage structure for workers at the lowest quantiles. However, this is not verified for workers at the upper half of the wage distribution. The main message from Figure 2 is that despite the fact workers had a definite improvement in their attributes (became more educated and acquired more experience in the labour market), the value placed by the labour market on



(a) Composition effect

(b) Structural effect

Figure 2: Counterfactual Analysis - aggregate decomposition

worker's characteristics has dropped generally, especially for higher-skilled workers.

6.2 Composition effects

Table 5 exhibits the results for the individual decomposition for covariates (left panel) and coefficients (right panel). Now, instead of constructing the counterfactual distribution of wages in 2018 if all covariates were distributed as in 1994 (x0b1), we construct two alternative counterfactual distributions: one if each single covariate was distributed as in 1994 ($f^*(w(1); y(0))$);, and other if each individual vector of coefficients was distributed as in 2018 ($f^*(w(0); \beta_y(1))$). Hence, the individual covariate and coefficient effect will be given by ($f^*(w(1)) - f^*(w(1); y(0))$) and ($f^*(w(0); \beta_y(1)) - f^*(w(0))$) respectively.

Starting from the left panel, we can observe that the covariates gender, tenure, and firm size had a negative impact on the wage distribution. There are mainly two reasons for this left location shift: 1) Regarding gender, as women earn less than men in all quantiles (as shown in Table 3), adding more women to the sample will shift the wage distribution to the left; 2) Regarding tenure and firm size, as the average tenure at the firm and firm size decreased from 1994 to 2018, these covariates had a negative impact on the wage distribution.

Contrarily to the aforementioned covariates, education and age had a positive impact on the wage distribution, and are quite more important in explaining changes in those distributions, especially education. As shown in Figure 3, the education composition effect almost fully accounts for the whole aggregate composition effect presented in Figure 2(a). Apart from this,

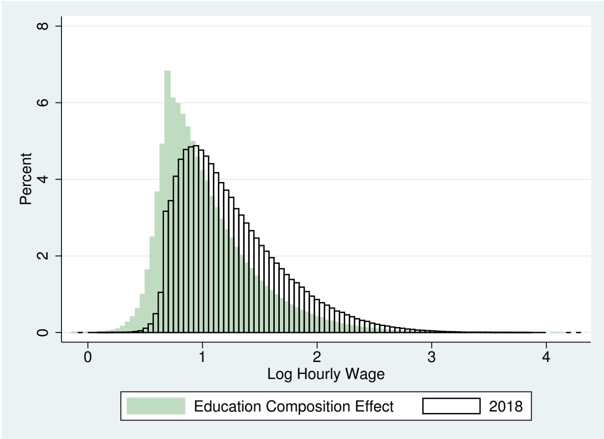


Figure 3: Counterfactual Analysis - individual covariates

what is most remarkable on the impact of age and education on wages is the monotonic growth of the estimated effects throughout the wage distribution: as we move from the 10th to the 90th quantile, the effect of education and age on wages (especially the first) keeps increasing, meaning wages at the top of the distribution have increased much more than wages at the bottom.

From Table 3, we have seen that the returns to education were monotonically increasing with the quantiles. By estimating wage distributions and counterfactuals, we can conciliate this evidence with changes that occurred in the composition of the workforce. In Portugal, the massive increase in education, the factor associated with higher within-group dispersion, meant the share of lower educated workers decreased and the share of more educated workers, the ones with more dispersed wage distributions, increased. Hence, changes in educational levels contributed to higher wage dispersion and greater inequality. Machado and Mata (2005) method offers an elegant solution to isolate the impact of education on wage inequality, and quantify it: holding coefficients constant, the change in the levels of education of the workforce would have contributed to an increase of 15.2 log points of the 90/10 quantile ratio. The education composition effect was not only the most important factor in explaining changes in the wage distribution, but also the most important one contributing to greater wage inequality.

The main message is that although there was a decrease in wage inequality from 1994 to 2018, and education is many times seen as an important tool to mitigate wage inequality, it actually was a driving force to a less equal wage distribution. Until now, we have considered that coefficients remained constant. However, the changes that occurred in the returns to education between 1994 and 2018 could also have contributed to greater wage inequality.

Table 5: Individual decomposition: covariates and coefficients

	Covariates					Coefficients					
Variables	Gender	Age	Tenure	Firm Size	Education	Gender	Age	Tenure	Firm Size	Education	Constant
10th quant.	-0.002***	0.026***	-0.000	-0.013***	0.136***	0.038***	-0.366***	-0.000	-0.132***	-0.079***	0.630***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
20th quant.	-0.004***	0.028***	-0.003***	-0.011***	0.159***	0.034***	-0.367***	0.008***	-0.132***	-0.077***	0.625***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
30th quant.	-0.006***	0.031***	-0.006***	-0.010***	0.179***	0.030***	-0.360***	0.016***	-0.130***	-0.073***	0.603***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
40th quant.	-0.008***	0.034***	-0.007***	-0.010***	0.197***	0.027***	-0.350***	0.024***	-0.128***	-0.071***	0.571***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
50th quant.	-0.011***	0.038***	-0.009***	-0.009***	0.216***	0.023***	-0.340***	0.034***	-0.125***	-0.067***	0.538***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
60th quant.	-0.013***	0.042***	-0.010***	-0.008***	0.235***	0.021***	-0.333***	0.045***	-0.124***	-0.062***	0.501***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
70th quant.	-0.015***	0.046***	-0.009***	-0.005***	0.256***	0.018***	-0.332***	0.056***	-0.124***	-0.057***	0.459***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
80th quant.	-0.018***	0.052***	-0.005***	-0.002***	0.276***	0.015***	-0.339***	0.068***	-0.127***	-0.050***	0.421***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
90th quant.	-0.023***	0.060***	0.004***	0.006***	0.288***	0.013***	-0.352***	0.085***	-0.127***	-0.033***	0.388***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90/10	-0.021***	0.035***	0.004***	0.018***	0.152***	-0.025***	0.014***	0.085***	0.004***	0.046***	-0.242***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90/50	-0.012***	0.023***	0.012***	0.015***	0.072***	-0.010***	-0.013***	0.051***	-0.002***	0.034***	-0.149***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
50/10	-0.009***	0.012***	-0.008***	0.004***	0.080***	-0.015***	0.026***	0.034***	0.007***	0.012***	-0.093***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)

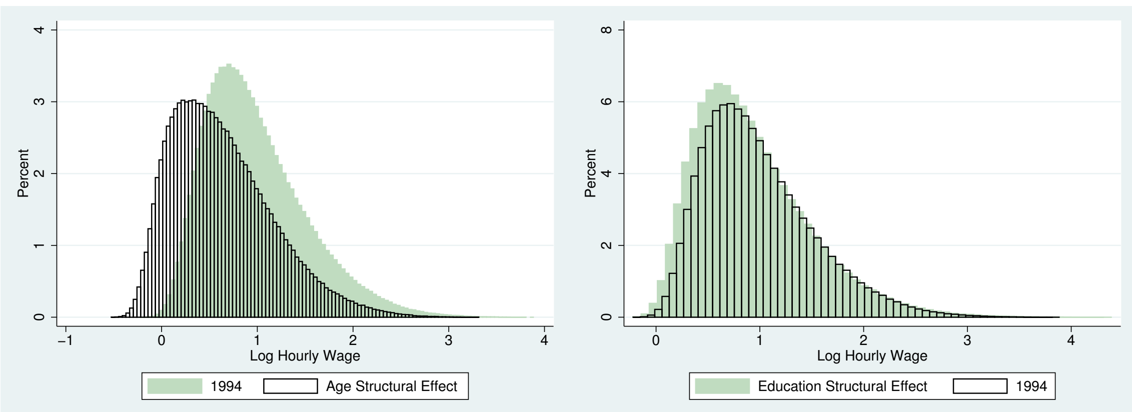
Notes: The left panel exhibits the contribution of individual covariates to the change in the selected quantiles of the wage distribution , comparing the wage distribution in 2018 with the counterfactual wage distribution in 2018 if each single covariate was distributed as in 1994.
The right panel exhibits the contribution of individual coefficients to changes in the wage distribution, obtained by comparing the counterfactual wage distribution in 1994 if single coefficients were distributed as in 2018 with the wage distribution in 1994.
Standard errors obtained from 100 bootstrap replications. **** $p<0.01$, ** $p<0.05$, * $p<0.1$

6.3 Structural effects

The right panel (Table 5) exhibits the individual effects of coefficients. Contrarily to the previous analysis, we are interested in how the coefficients changed for each quantile between 1994 and 2018, not in changes in the composition of the workforce. As we can observe, only the changes in coefficients of gender and tenure have contributed to the right shift of the wage distribution, since 1) the gender wage gap decreased for all quantiles; 2) the return to tenure has increased for all quantiles, except for the 10th quantile. The distribution of the returns to tenure has become more dispersed since top quantiles benefited much more from higher returns.

An interesting feature reported in Table 5 (with more detail in Figure 4(a)) is the decrease in the returns to age and firm size, especially the first one. As Portugal, Raposo, and Reis (2018) reported, there was a steep decline in the value placed in labour market experience in Portugal, especially for workers located at the lower quantiles. This drop had a major influence in explaining the negative contribution of all coefficients to the aggregate decomposition.

Finally, it can be seen that the observed decline in the returns to education shifted the wage distribution to the left. However, this decrease was more pronounced and concentrated at the lower part of the wage distribution. If we consider a higher number of percentiles (Figure 4(b)), we observe that the value placed on education for workers at top quantiles did not decrease. Overall, this implies that the change in the distribution of returns to education in itself would have increased the 90/10 ratio by 4.6 log points (holding composition effects constant), this increase being essentially concentrated at the higher half of the wage distribution.



(a) Age Structural effect

(b) Education Structural effect

Figure 4: Counterfactual Analysis - individual coefficients

6.4 Robustness check

Until now, the variable education has only been measured in a continuous fashion. By using years of schooling as a proxy for education, we concluded that more educated individuals have more dispersed wage distributions. However, as explored by Lemieux (2006a), the amount of within-group dispersion among college-educated workers is much higher than among lower education groups. Hence, to examine whether within-group dispersion is particularly higher for this group, I also applied a different specification from (2), converting education into a discrete variable, that has a value 1 for college graduates and 0 otherwise. With this specification, we can assess the college/non-college wage differential for the selected quantiles.

Table 7 in the Appendix shows the results for the individual decomposition under the new specification. The reported results for the composition effects are very similar to the ones in Table 5. The main difference concerns the education composition effect, which is much more inequality increasing than before. The increase in the share of college graduates in the workforce was responsible for an increase of 19.7 log points of the 90/10 ratio. Furthermore, much of this increase was concentrated in the upper half of the wage distribution, a different finding from the results reported in Table 5.

On the other hand, the results for the structural effects are somewhat different relative to previous results. Under this new specification, from 1994 to 2018, the returns to experience and firm size have decreased much more in the higher quantiles. Additionally, the change in returns to tenure had a greater impact on wage inequality. Finally, one major difference concerns the decrease in the returns of having a college degree for all quantiles, this decrease being slightly more pronounced in the higher quantiles. Combining the education composition and structural effects under this new specification, we conclude that individuals with a college degree have indeed more dispersed wage distributions than lower educated individuals, but from 1994 to 2018, the amount of within-group inequality among this group has slightly decreased.

7 Final Remarks

7.1 Discussion

The results shown in this work present evidence that, from 1994 to 2018, education contributed unequivocally to more wage inequality in Portugal, counteracting opposing forces that contributed to less overall wage inequality. Comparing these results with the results obtained in Machado and Mata (2005), I detect a stronger impact of the education composition effect on wage inequality and explore the role of changing returns to education in explaining inequality changes. Overall, education has contributed to higher wage inequality via two mechanisms: composition effects (observed increase in educational levels contributed to a more unequal wage distribution) and structural effects (observed changes in returns to education affected much more individuals at the bottom of the wage distribution). The two factors alone contributed to an increase of 15.2 and 4.6 log points in the 90/10 ratio, respectively.

Conventional wisdom predicts that increasing the educational levels of the workforce translates into a more balanced wage distribution. The problem with this claim is that, as this study shows, the returns to education are not equally shared along the wage distribution: education increases much more wages of individuals at the top of the wage distribution than at the bottom. In other words, more educated individuals have more dispersed wage distributions. Given the fact this tendency has been aggravated from 1994 to 2018, by increasing the share of more educated individuals in the economy, education is contributing to greater wage inequality. The main question still left unanswered is why do more educated individuals have more spread out wage distributions? I propose three different explanations.

Firstly, there may exist some unobserved information about the education of an individual. Despite the fact the variable education in our model captures differences in years of schooling or the maximum schooling level obtained, it does not capture differences in school quality or areas of study, two unobserved attributes that could contribute to more within-group dispersion. It could be that individuals at the bottom of the conditional wage distribution, having the same "amount" of education relative to individuals located at the top of the same distribution, have studied in worst schools or have chosen areas of study with lower labour market returns. These unobserved attributes are more dispersed among workers with tertiary education rather than secondary education since there is much more heterogeneity of choice in opting for a college degree or university than at the high school level. Table 8 in the Appendix shows the regular wages (monthly basis) by area of study for a college graduate from a sample of 393,061 individuals in 2018. As we can see, the labour market places much more value on quantitative areas of study (Mathematics, Engineering, and IT) oriented to the use of new technologies, and less value on areas related to Arts, Teaching, and Social Services. Regarding school quality, better and more selective universities may attract the brightest students and offer better training in specific areas and better access to job opportunities in the labour market.

A second plausible explanation is over-education. If higher educated individuals end up performing tasks that require fewer skills (and consequently worst paid jobs), the same workers would be at a lower position in the wage distribution, exhibiting lower returns to education and contributing to more within-group dispersion. Given the prominent rise in the proportion of college graduates in Portugal, we might expect to see an increase in the share of overeducated individuals with a college degree if the demand for more educated employees does not keep up with supply. Pimenta and Pereira (2019) report that this share increased from 1995 to 2005 (from 20 to 30%) and stabilized until 2013. Despite this increase, the authors consider that the majority of incoming college graduates in this period end up performing high skill tasks.

Finally, an alternative explanation concerns ability. Using a high-dimensional fixed effects model, Cardoso, Guimarães, Portugal, and Reis (2018) report that for Portugal, the dispersion of worker abilities, captured by the worker fixed effects, is much higher for college graduates than for individuals with lower education. Additionally, the wage differential between the most and less able workers should be considerably higher for jobs that require higher skills and more abstract tasks (and consequently more education) than for medium/low-skilled jobs that require routine or manual tasks. These two factors combined explain why more educated individuals have indeed more disperse wage distributions.

7.2 Conclusion

Concluding, this paper assesses the impact of education on changes in the Portuguese wage distribution, from 1994 to 2018. During this period, the overall level of wages increased (real average hourly wages grew by 25%), meaning the wage distribution shifted to the right. However, this growth was more pronounced in the bottom quantiles of the wage distribution, leading to a decline in wage inequality measures. The presented results show that 1) education was the most decisive covariate contributing to the overall wage growth; 2) the covariate that most contributed to greater wage inequality, mostly for two reasons. First, the returns to education are much higher in the upper half of the wage distribution than in the lower half, implying that an increase in the overall educational levels of the economy would by itself contribute to higher wage dispersion. Second, the returns to education became even more dispersed between 1994 and 2018.

This paper does not explore further the implications of the reported results. However, a more cautious analysis should be done in the aftermath of this work to understand the real mechanisms lying behind the inequalities in the graduate's remunerations. While most educational policies are oriented to promote widespread access to College Education, they should also be focused on correcting the existing inequalities in the College Education system, answering questions such as: which specific college degrees serve as a channel to higher-paying jobs; which degrees have a better match between areas of study and market needs (job titles); and which degrees prepare individuals to perform abstract tasks that are more difficultly replaced by new disruptive technologies.

8 References

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2005. ”Rising Wage Inequality: The Role of Composition and Prices.” *NBER Working Paper No. 11628.*

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. ”Trends in U.S. wage inequality: Revising the revisionists.” *Review of Economics and Statistics*, 90(2), 300–323.

Blinder, Alan. 1973. ”Wage Discrimination: Reduced Forms and Structural Estimation.” *Journal of Human Resources*, 8(4), 436–455.

Campos, Maria and Hugo Reis. 2017. “Uma Reavaliação do Retorno do Investimento em Educação na Economia Portuguesa.” *Revista de Estudos Económicos Banco de Portugal*, 3(2), 1–30.

Cardoso, Ana R., Paulo Guimarães, Pedro Portugal, and Pedro Raposo. 2016. ”The sources of the gender wage gap.” *Bank of Portugal Economic Studies*, 2(2),47–66.

Cardoso, Ana R., Paulo Guimarães, Pedro Portugal, and Hugo Reis. 2018. ”The returns to schooling unveiled.” *IZA Discussion Papers No. 11419*.

Centeno, Mário and Álvaro A. Novo. 2009. ”When supply meets demand: wage inequality in Portugal.” *IZA Discussion Papers No. 4592*.

DiNardo, John E., Nicole Fortin, and Thomas Lemieux. 1996, ”Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach.” *Econometrica*, 64(5), 1001–1044.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. ”Decomposition Methods in Economics.” *Handbook of Labor Economics*. Edited by Orley Ashenfelter and David Card, 4(1), 1–102.

Goos, Maarten, and Alan Manning. 2007. ”Lousy and lovely jobs: The rising polarization of work in Britain.” *The Review of Economics and Statistics*, 89(1), 118–133.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. “Wage Inequality and the Rise in Returns to Skill.” *Journal of Political Economy*, 101(3), 410–442.

Koenker, Roger and Gilbert Bassett, Jr. 1978. “Regression Quantiles.” *Econometrica*, 46(1), 33–50.

Lemieux, Thomas. 2002. “Decomposing Changes in Wage Distributions: A Unified Approach.” *Canadian Journal of Economics*, 35(4), 646–688.

Lemieux, Thomas. 2005. “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” University of British Columbia.

Lemieux, Thomas. 2006a. ”Post-secondary Education and Increasing Wage Inequality.” *American Economic Review*, 96(2), 195–199.

Lemieux, Thomas. 2006b. ”Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” *American Economic Review*, 96(3), 461–498.

Machado, José A. F. and José Mata. 2005. “Counterfactual decomposition of changes in wage distributions using quantile regression.” *Journal of Applied Econometrics*, 20(4), 445–465.

Mincer, Jacob A. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research, Inc.

Mincer, Jacob A. 1997. ”Changes in Wage Inequality, 1970-1990.” *Research in Labor Economics*, 16, 1–18.

Oaxaca, Ronald. 1973. “Male-Female Wage Differentials in Urban Labor Markets.” *International Economic Review*, 14(3), 693–709.

Portugal, Pedro, Pedro Raposo, and Hugo Reis. 2018. ”The distribution of wages and wage inequality.” *Bank of Portugal Economic Studies*, 4(1), 1–14.

Pimenta Ana C. and Manuel C. Pereira. 2019. ”Aggregate educational mismatches in the Portuguese labour market.” *Bank of Portugal Economic Studies*, 5(1), 41–66.

9 Appendix

Table 6: Within-Group Dispersion for Education and Age Categories

		Inequality Measures		Worker Share	
		1994	2018	1994	2018
Less than 9th grade					
	90/10	0,786	0,543		
Age<36	90/50	0.493	0.354	0.333	0.022
	50/10	0.293	0.189		
	90/10	1.057	0.623		
Age 36-45	90/50	0.655	0.445	0.172	0.051
	50/10	0.402	0.178		
	90/10	1.145	0.694		
Age>45	90/50	0.681	0.491	0.169	0.141
	50/10	0.464	0.203		
	90/10	1.038	0.668		
All	90/50	0.658	0.470	0.674	0.214
	50/10	0.380	0.198		
9th grade/High School					
	90/10	1.305	0.622		
Age<36	90/50	0.764	0.420	0.176	0.218
	50/10	0.541	0.202		
	90/10	1.331	0.938		
Age 36-45	90/50	0.575	0.653	0.063	0.174
	50/10	0.756	0.285		
	90/10	1.491	1.250		
Age>45	90/50	0.648	0.914	0.040	0.171
	50/10	0.843	0.336		
	90/10	1.483	0.940		
All	90/50	0.816	0.690	0.279	0.563
	50/10	0.667	0.250		
College degree					
	90/10	1.423	1.065		
Age<36	90/50	0.616	0.567	0.024	0.096
	50/10	0.807	0.498		
	90/10	1.444	1.431		
Age 36-45	90/50	0.612	0.725	0.013	0.081
	50/10	0.832	0.706		
	90/10	1.662	1.748		
Age>45	90/50	0.687	0.774	0.009	0.044
	50/10	0.975	0.974		
	90/10	1.566	1.491		
All	90/50	0.679	0.844	0.046	0.221
	50/10	0.887	0.642		

Table 7: Individual decomposition: covariates and coefficients-College/Non-College

	Covariates					Coefficients					
Variables	Gender	Age	Tenure	Firm Size	College	Gender	Age	Tenure	Firm Size	College	Constant
10th quant.	-0.004***	0.003***	-0.006***	-0.011***	0.014***	0.038***	-0.306***	0.000***	-0.130***	-0.000***	0.758***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
20th quant.	-0.004***	0.005***	-0.006***	-0.011***	0.024***	0.033***	-0.339***	0.006***	-0.141***	-0.001***	0.749***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
30th quant.	-0.006***	0.007***	-0.006***	-0.012***	0.034***	0.028***	-0.373***	0.011***	-0.149***	-0.002***	0.749***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
40th quant.	-0.007***	0.010***	-0.005***	-0.012***	0.049***	0.022***	-0.410***	0.018***	-0.157***	-0.003***	0.759***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
50th quant.	-0.008***	0.013***	-0.004***	-0.010***	0.070***	0.017***	-0.453***	0.029***	-0.165***	-0.005***	0.783***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
60th quant.	-0.010***	0.019***	-0.002***	-0.008***	0.099***	0.012***	-0.504***	0.045***	-0.175***	-0.008***	0.818***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
70th quant.	-0.013***	0.025***	0.001	-0.004***	0.136***	0.008***	-0.565***	0.068***	-0.187***	-0.014***	0.864***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
80th quant.	-0.019***	0.033***	0.004***	0.001	0.177***	0.005***	-0.635***	0.099***	-0.203***	-0.023***	0.913***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
90th quant.	-0.025***	0.043***	0.005***	0.006***	0.211***	0.002***	-0.659***	0.145***	-0.222***	-0.045***	0.962***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90/10	-0.021***	0.041***	0.011***	0.016***	0.197***	-0.036***	-0.353***	0.145***	-0.091***	-0.045***	0.204***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90/50	-0.017***	0.030***	0.009***	0.016***	0.141***	-0.015***	-0.205***	0.116***	-0.057***	-0.040***	0.179***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
50/10	-0.005***	0.011***	0.002***	0.000	0.056***	-0.021***	-0.148***	0.028***	-0.035***	-0.005***	0.025***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: The presented results refer to the individual decomposition using education as a categorical variable that reports the value 1 if an individual has a college degree and 0 otherwise. Standard errors obtained from 100 bootstrap replications. **** p<0.01, ** p<0.05, * p<0.1

Table 8: Regular Wage by Area of Study of College Degree-2018

		Sample	
Area of Study	Regular wage (€)	%	N
Transport & Logistics	2891.336	0.12	473
Security Services	2541.418	0.17	653
Mathematics and Statistics	2303.911	1.07	4,216
Law	2301.811	2.24	8,823
Engineering	2247.488	14.27	56,094
IT	2106.849	4.62	18,140
Manufacturing	2101.653	0.27	1,067
Business	2084.006	17.78	69,891
Physical Sciences	1949.033	0.90	3,538
Behavioural & Social Sciences	1942.588	6.41	25,212
Architecture & construction	1791.689	1.67	6,567
Health	1791.532	11.48	45,122
Agriculture & Fishery	1773.549	0.75	2,966
Life Sciences	1697.606	1.44	5,649
Information & Journalism	1600.388	1.63	6,397
Humanities	1599.454	2.86	11,234
Environment Protection	1545.473	0.32	1,255
Unknown / not reported	1517.554	23.61	92,799
Personal Services	1481.309	0.55	2,166
Teaching and Education	1405.855	3.94	15,475
Veterinary Sciences	1397.952	0.48	1,876
Arts	1286.163	1.51	5,950
Social Services	1256.584	1.91	7,498
Total		100	393,061

Notes: Values obtained from Quadros de Pessoal- year 2018. Restricts to full time wage earners with a college degree.